Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

Claim 1. (currently amended) A system for annotating a group of subsets of genes with words and phrases that characterize each individual subset of genes and that also distinguish said individual subset of genes from each and every of the other individually considered subsets of genes, by displaying words and phrases taken from literature abstracts and other text corresponding to each subset of genes, arranged in the order of sorted numerical weights that the system assigns to the words and phrases, comprising:

(a) means for identifying a set of genes;

(b) means for partitioning the set of genes in (a) into disjoint subsets of genes known as clusters;

(c) means for associating a set of literature documents with each gene in the set of genes in (a), and a means for receiving the text of part or all of each said literature document;

(d) means for constructing a compendium of text for each individual subset of genes known in (b) as a cluster, said compendium consisting of the text received for all of the literature documents that had been associated and received in (c), for all genes that are members of the subset of genes in said cluster;

(e) [[means for assigning numerical weights to words or phrases contained in the compendia of text constructed in (d), said assignment being made by the application, to the words or phrases in those compendia of text constructed in (d), of any of the word weight-setting methods that are implemented in the computer program Rainbow, said application being intended to annotate]] means for running a publicly available computer program known as Rainbow, the running of which is used as a means for annotating the group of subsets of genes known as clusters in (b), with words and phrases that characterize each individual subset of genes known as a cluster in (b) and that also distinguish said individual subset of genes from each and every of the other individually considered subsets of genes known as a cluster in (b), in terms of different words and phrases that the system attaches to different individual subsets of genes, each of which is known as a cluster in (b) [[;]] , said annotating being performed in the following two stages:

-- first, instruct the computer program Rainbow to take as extrinsic input-data each and every

compendium of text that was constructed in (d) for each individual subset of genes known in (b) as a cluster, then process those input-data to produce as output-data a statistical model of the text in all those compendia; and

-- second, instruct the computer program Rainbow to take as its input-data the aforementioned statistical model of the text, and to process that data to produce as output-data a list of words and phrases for each subset of genes known in (b) as a cluster, along with word-weights that Rainbow calculates for each word or phrase in the list, the magnitude of which indicates the weight that the system attaches to the corresponding word or phrase as a characterization of the subset of genes known in (b) as a cluster, said word-weights being calculated by default through Rainbow's implementation of the Naive Bayes algorithm, or optionally through Rainbow's implementation of other word weight-setting algorithms.

(f) means for sorting, pruning, storing, and displaying the words and phrases contained in a compendium of text [[associated with]] constructed in (d), for each of the individual subsets of genes known as clusters in (b), said sorting being based on the magnitude of the numerical weights assigned to the corresponding words and phrases as provided in (e) for each subset of genes known as a cluster in (b); said pruning being based on the setting of a minimum cutoff for the magnitude of the numerical weights assigned to the corresponding words and phrases as provided in (e) for each subset of genes known as a cluster in (b); [[,]] and said storage and display allowing words and phrases, for each subset of genes known as a cluster in (b), to be arranged in the descending order of their [[sorted]] corresponding numerical weights;

whereby the words and phrases having the greatest numerical weights for each individual subset of genes known as a cluster in (b) provide an indication of the concepts, structures, functions, and processes with which said individual subset of genes known as a cluster in (b) is most particularly associated, and with which said individual subset of genes is also distinguished from each and every of the other individually considered subsets of genes known as clusters in (b), in terms of different words and phrases that the system attaches to different individual subsets of genes known as clusters in (b).

Claim 2. (canceled) A system for evaluating the quality of gene clustering, comprising:

(a) means for identifying a set of genes;

(b) means for partitioning the set of genes in (a) into subsets known as clusters;

(c) means for associating a set of documents with each gene in the set of genes in (a), and consequently a means for associating a set of documents with each of the clusters in (b);

(d) means for partitioning the set of documents in (c) into two subsets, a training subset and a testing subset;

(e) means for receiving the text of part or all of each of the training subset documents in (d);

(f) means for receiving the text of part or all of each of the testing subset documents in (d);

(g) means for using words or phrases in the text of documents in (e) to train a document classifier, said training being accomplished by partitioning the documents according to their association with each cluster as provided in (b) and (c), followed by the parameter-fitting, using the words or phrases in those partitioned documents, of any of the document classifiers that are implemented in the computer program Rainbow;

(h) means for using words or phrases in the text of each document in (f) to test the trained document classifier in (g), wherein the classifier predicts the cluster with which the test document is associated;

(i) means for the option of calculating and storing the fractions of test documents in (d) known to correspond to each cluster as provided in (b) and (c), that are correctly predicted to be associated with each cluster, upon testing with the document classifier as provided in (h);

(j) means for the option of repeatedly and randomly partitioning documents in (c) into training and test subsets as provided in (d), for using each such partitioning to calculate a fraction of correct classifications for each cluster as provided in (e)-(i), and for storing said fractions for each and every such random partitioning of

documents into training and test subsets.

(k) means for the option of repeatedly and randomly partitioning the set of genes in (a) into subsets, wherein the sizes of the random subsets are matched to the sizes of the clusters as provided in (b); for re-associating a set of documents with each gene in the set as in (c), and consequently associating a set of documents with each of the randomly partitioned subsets of genes; for making available means (d)-(i) so as to be able to calculate a fraction of correct classifications for each of the random partitions that are matched to the clusters as provided in (b); and for storing said fractions for each and every such random partitioning of the set of genes in (a).

(l) means for the option of calculating a measure of central tendency, such as mean or median, for the fractions that were generated by repeated, random partitioning of documents in (j), and for the fractions that were generated by repeated, random partitioning of the set of genes in (k); and for calculating a figure-of-merit for each cluster as the numerical difference between said measure of central tendency obtained from (j) and (k);

whereby said figure-of-merit for each cluster provides an indication of the extent to which some words and phrases, present in documents associated with genes in said cluster, collectively distinguish that cluster from all the other clusters, and whereby said figure-of-merit for each cluster provides an indication of the extent to which the annotations produced by the system of Claim 1 distinguish the clusters, and whereby said figure-of-merit for each cluster provides an indication of the quality of that cluster.

Claim 3. (canceled) A system for evaluating the quality of gene clustering, comprising:
(a) means for identifying a set of genes;
(b) means for partitioning the set of genes in (a) into subsets known as clusters;
(c) means for associating a set of documents with each gene in the set of genes in (a);

(d) means for calculating for every pair of genes within a cluster in (b) a coupling strength index, said index being proportional to the number of times that any document in (c) is associated with both members of said pair of genes; and for storing said set of index values for every cluster;

(e) means for repeatedly and randomly partitioning the set of genes in (a) into subsets, wherein the sizes of the random subsets are matched to the sizes of the clusters as provided in (b); for re-associating a set of documents with each gene in the set as in (c); for making available means (d) so as to be able to calculate coupling strength indices; and for storing said set of index values for every such random subset;

(f) means for calculating for every cluster in (b) and random subset in (e) a measure of central tendency, such as the mean or median, for the set of coupling strength index values that are stored as provided in (d) and (e);

(g) means for calculating and displaying for every cluster in (b) the percentage of times that the central tendency calculated in (f) is larger than the corresponding central tendency in (f), among those calculated repeatedly for corresponding random subsets in (e).

whereby said percentage for each cluster provides an indication of the extent to which documents associated with genes in said cluster collectively distinguish that cluster from all the other clusters, and whereby said percentage for each cluster provides an indication of the quality of that cluster.

Claim 4. (canceled) A method, in a computer system, of clustering gene expression data,
  wherein data received for each one of a plurality of genes constitute the
    response of said gene to an intervention at an initial time; and
  wherein data received for each one of a plurality of said genes were collected
    at a series of time points following the intervention; and
  wherein data received for a plurality of said genes are proportional to the amount
    of messenger RNA, x, for each one of said genes; and
  wherein the rate of change of x, dx/dt, is represented as a synthesis rate, f,

minus the product of a degradation rate k and x, namely, -kx; and

wherein the f and k are both represented as being piecewise continuous,

time-varying functions at each of the measurement time points for each

one of said plurality of genes; and

wherein f and k are approximated as truncated Taylor series, the coefficients of which

are estimated using the received data x, at each of the measurement time points

for each one of said plurality of genes; and

wherein the estimated values of the synthesis rate, f, at each of the measurement

time points for each one of said plurality of genes are used to cluster said plurality

of genes;

whereby said clustering may reveal subsets of genes among the plurality of genes

that are regulated by the same transcription factors, as evidenced by the similarity

of their time-varying transcription rates.

Claim 5. (canceled) The method of claim 4, used as the means for partitioning a set

of genes into clusters in a system for annotating sets and subsets of genes.

Claim 6.(canceled) The method of claim 4, used as the means for partitioning a set

of genes into clusters in a system for evaluating the quality of gene clustering.